Estimating Time From Atomic Clocks

Richard H. Jones* and Peter V. Tryon**

National Bureau of Standards
Boulder, CO 80303

A Kalman recursive algorithm for estimating time from an ensemble of atomic clocks has been developed. The algorithm allows for the addition or deletion of clocks at any time, and provides automatic error detection and correction. The observations consist of time differences between clocks and may be taken at unequally spaced time points. Maximum likelihood estimates of the unknown parameters are obtained with confidence intervals, as well as hypothesis tests to determine whether the estimated parameters are significantly different from zero. The program is operational on the National Bureau of Standards' Time and Frequency Division's PDP 11/70.

Key words: atomic clocks; Kalman recursion; maximum likelihood estimation; missing observations; nonlinear estimation; state space; time series analysis; unequally spaced data.

1. Introduction

Cesium beam atomic clocks have an accuracy of a few parts in 10^{14} over a period of a day; however, they are not deterministic and undergo stochastic variations in both time and frequency. It was shown by Tryon and Jones [1] that the actual frequency of a clock behaves as a random walk which, over a time period of a day, has a standard deviation of less than four nanoseconds per day. The effect of this random walk in frequency on time measurements is that the individual frequency steps are summed, producing an integrated random walk.

In addition, frequency has a white noise component which integrates into an independent random walk in time with a standard deviation of up to 15 nanoseconds over a time period of a day. Therefore, the deviation of clock time from true time behaves as a random walk plus the sum of a random walk. The possibility of a frequency drift also exists.

When estimating time from an ensemble of clocks, the only observations possible are measurements of time differences between clocks. In addition, observations may occur at unequally spaced time points, and various types of errors are possible. The most common errors are *read errors*, where a single measurement is incorrect but subsequent measurements are the same as if the read error did

not occur; *time steps*, where a jump in time occurs and the clock remains at the new position; and *frequency steps*, where the frequency takes a step and remains at the new position. This last error appears as a change in the rate of gain of the clock.

This paper describes a state space algorithm for estimating time from an ensemble of cesium beam atomic clocks with unequally spaced observations subject to various errors. Clocks can be added to or deleted from the ensemble. Using iterative calculations over a recent history of the ensemble, maximum likelihood estimates of the unknown variances can be obtained by nonlinear optimization. Confidence intervals on the estimated parameters and test of hypotheses, such as whether parameters are significantly different from zero, can also be obtained from the likelihood function.

2. Mathematical Model

The model for a single clock in state space form is

$$
\begin{bmatrix} x(t) \\ y(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} 1 & \delta(t) & \delta^2(t)/2 \\ 0 & 1 & \delta(t) \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(t-1) \\ y(t-1) \\ w(t-1) \end{bmatrix} + \begin{bmatrix} \varepsilon(t) \\ \eta(t) \\ \sigma(t) \end{bmatrix} \quad (2.1)
$$

where $t = 1, 2, 3, \ldots$ is an index of the observation points, $x(t)$ is the difference between the clock's time and "true" time, and $y(t)$ is the difference between the clock's frequency and the fundamental resonance of the cesium atom which defines the second, 9,192,631,770 Hz, and

---
*Division of Biometrics, Box B-119, School of Medicine, University of Colorado, Denver, CO 80262.
**Deceased. Dr. Tryon served with the Center for Applied Mathematics, National Engineering Laboratory.

is expressed in units of nanoseconds per day. $w(t)$ represents a possible frequency drift in units of nanoseconds/day^2, and $d(t)$ is the time interval between $t-1$ and t in days. $\epsilon(t)$, $\eta(t)$ and $\alpha(t)$ are random variables with zero mean, uncorrelated with each other, and uncorrelated in time. These are the input to the random walks, and for small $d(t)$, their variances are proportional to the length of the time intervals between observations, $d(t)$,

$$\text{Var } \{\epsilon(t)\} = d(t)\, \sigma_\epsilon^2$$

$$\text{Var } \{\eta(t)\} = d(t)\, \sigma_\eta^2 \qquad (2.2)$$

$$\text{Var } \{\alpha(t)\} = d(t)\, \sigma_\alpha^2$$

If $\sigma_\alpha^2 = 0$, the drift, $w(t)$, does not change with time.

The state equation for an ensemble of m clocks is obtained by concatenating the three state elements of each clock into a column vector of length 3m. The state transition matrix, $\Phi(t)$, is a 3m by 3m matrix consisting of 3 by 3 blocks on the main diagonal corresponding to the state transition matrix in eq (1), and zero blocks off the main diagonal. This equation of state for m clocks will be written.

$$X(t) = \Phi(t)\ X(t-1) + U(t). \qquad (2.3)$$

The random input vector, $U(t)$, is of length 3m and consists of the $\epsilon(t)$, $\eta(t)$, and $\alpha(t)$ for each clock. The covariance matrix of $U(t)$, $d(t)$ Q, is a 3m by 3m diagonal matrix with diagonal elements consisting of the σ_ϵ^2, σ_η^2 and σ_α^2 for each clock. While reasonable guesses are available for the σ_ϵ^2 and σ_η^2, it is necessary to obtain better estimates of these variances from data. The values are characteristics of each clock and will be different for each clock. It is also necessary to determine whether the σ_α^2 are significantly different from zero. If the σ_α^2 are not significantly different from zero it would indicate that any drifts that exist are deterministic rather than random walks. In this case the $w(t)$ will be constant with respect to time and these can be tested to determine if the drifts are significantly different from zero.

When observations are taken, one of the clocks is used as a reference clock, and the time differences between this clock and each of the other clocks are recorded. This gives an observation vector $Z(t)$, of length $m-1$, and an observation equation of the form

$$Z(t) = H(t)X(t) + V(t), \qquad (2.4)$$

where $H(t)$ is a matrix indicating which clock differences are observed. If the first clock is the reference clock, $H(t)$

is of the form

$$H(t) = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \end{bmatrix} \qquad (2.5)$$

and has dimension $(m-1)$ by $3m$. $H(t)$ can be time dependent since if observations are missing or deleted because of error detection, rows of $H(t)$ are eliminated. $V(t)$ is a vector of observational errors. These errors are very small when reading clock differences. If the data are truncated to the nearest nanosecond, the variances of the elements $V(t)$ would be

$$r = 1/12 \text{ (nanoseconds)}^2. \qquad (2.6)$$

When observations are taken with higher precision than the nearest nanosecond, r can be reduced and its value determined from the instrumentation. The observational error variance matrix, R, will be a diagonal matrix and diagonal elements r, assumed to be equal and known. While these numbers are very small, the inclusion of R in the recursion does add numerical stability.

3. The Data

In a companion paper [1] Tryon and Jones used 333 daily observations from seven clocks starting February 16, 1979 to estimate clock parameters. These data were chosen since they consisted of a fairly long record with few errors. Some preprocessing was necessary to remove several outliers giving a set of data that could be considered error-free. These calculations were carried out on a CDC Cyber 170/750 computer.

This paper reports the new statistical procedures and techniques that were incorporated when the algorithm was rewritten to run on the National Bureau of Standards Time and Frequency Division's PDP-11/70. The algorithm was initialized on March 31, 1981 with the first set of data collected on April 1, 1981. At the beginning of each month the daily observations from the previous month are passed to the algorithm without preprocessing. While the data are nominally collected at the same time each day, for various reasons, the time of the observations are sometimes off by several hours giving unequally spaced data. Individual clocks may have read errors, time steps, and frequency steps. In addition, clocks may be added or deleted from the ensemble.

18

4. The Kalman Recursion With Error Detection

Let $X(t|s)$, $s \leq t$ be the best estimate of the state vector at time t given observations up to time s, and let $P(t|s)$ be its covariance matrix. The recursion begins by specifying $X(0|0)$ and $P(0|0)$. First a one step prediction of the state vector is calculated [3]

$$X(t+1|t) = \Phi(t+1) \, X(t|t), \qquad (4.1)$$

as is its covariance matrix

$$P(t+1|t) = \Phi(t+1) \, P(t|t) \, \Phi'(t+1) + \delta(t+1)Q. \quad (4.2)$$

The predicted values of the observations at time $t+1$ are

$$Z(t+1|t) = H(t+1) \, X(t+1|t). \qquad (4.3)$$

The difference between the vector of actual observations and predicted observations is the *innovation* vector,

$$I(t+1) = Z(t+1) - Z(t+1|t), \qquad (4.4)$$

which has covariance matrix

$$C(t+1) = H(t+1)P(t+1|t) \, H'(t+1) + R(t+1). (4.5)$$

It is at this point in the recursion that a statistical method of error detection is employed. A simple method would be to divide each element of the innovation vector by the square root of corresponding diagonal element of its covariance matrix giving a standard normal variable under the null hypothesis of no error. However, since clock differences are being measured and the elements of the innovation vector are intercorrelated, this is not an optimal test.

If there is an error in the reference clock, a constant bias will appear in every reading. This can be written as a regression equation with correlated errors

$$I(t+1) = A_1 \, \beta + \epsilon, \qquad (4.6)$$

where A_1 is a column vector of ones

$$A_1' = [1, 1, \ldots 1], \qquad (4.7)$$

and the error vector, ϵ, has covariance matrix $C(t+1)$ from eq (4.5). If there is an error in a clock which is not the reference clock, the error will appear only in the measurement involving that clock. For clock i, the model is

$$I(t+1) = A_i \, \beta + \epsilon, \qquad (4.8)$$

where

$$A_i' = [0 \; 0 \; \cdots -1 \cdots 0], \qquad (4.9)$$

the minus one appearing in the position of the clock. Although there are only $m-1$ measurements, m tests, one for each clock, are possible, but they are not statistically independent tests. Letting A represent any of the A vectors, the minimum variance weighted least squares estimate of β is

$$b = A'C^{-1}(t+1)I(t+1)/A'C^{-1}(t+1)A, \qquad (4.10)$$

and has standard error

$$\text{s.e.}(b) = (1/A'C^{-1}(t+1)A)^{\frac{1}{2}}. \qquad (4.11)$$

If the test statistic

$$z = b/\text{s.e.}(b) \qquad (4.12)$$

has absolute value larger than some value such as 3.0, the corresponding clock is assumed to have an error. It must be remembered that this test is only approximate if guesses are being used for the random walk variances in the Q matrix, and will be much more accurate after maximum likelihood estimates are obtained for these variances. The details of the above calculations are given in the Appendix.

An overall test for errors is possible by calculating the quadratic form

$$Q = I'(t+1) \, C^{-1}(t+1) \, I(t+1) \qquad (4.13)$$

which will be distributed as chi-square with $m-1$ degrees of freedom under the null hypothesis of no errors in the observations. Although this statistic is calculated, it is no longer being used for error detection since it is felt that this test lacks power against the most common alternative of an error in a single clock.

If an error is detected, the reading for the clock in question is eliminated from the observation vector. If more than one error is detected the clock with the largest absolute value of z is eliminated and the tests recalculated for the remaining clocks. If the clock to be eliminated is the reference clock, the reference clock is eliminated by choosing a clock still in the model to be the new reference clock and forming a new data vector by subtracting the reading of the difference between the previous reference clock and the new reference clock from each of the other readings. This is shown schematically as:

19

Previous   New

$①-②$

$①-③-\begin{bmatrix}①-②\end{bmatrix}=②-③$

$①-④-\begin{bmatrix}①-②\end{bmatrix}=②-④$ (4.14)

$①-⑤-\begin{bmatrix}①-②\end{bmatrix}=②-⑤.$

When all remaining clocks pass the test, the recursion continues. Let the Kalman gain be

$$\Delta(t+1) = P(t+1|t)\, H'(t+1)\, C^{-1}(t+1). \quad (4.15)$$

If clocks have been eliminated due to errors, the rows of $H(t+1)$ corresponding to the eliminated clocks are eliminated as are the corresponding rows and columns of $C(t+1)$. In the cases where the reference clock is changed, the positions of the ones in the $H(t+1)$ matrix must be moved to correspond to the new reference clock. The estimated state vector is then updated,

$$X(t+1|t+1) = X(t+1|t) + \Delta(t+1)\, I(t+1), \quad (4.16)$$

as is its covariance matrix

$$P(t+1|t+1) = P(t+1|t) - \\ \Delta(t+1)\, H(t+1)\, P(t+1|t). \quad (4.17)$$

This completes the recursion, but the effects of any detected errors must be resolved. Because the measurement errors of the data are so small, the final updating eq (4.16) essentially sets the time states to agree with the actual readings of the differences. This will not be the case for any clock that was eliminated at this step since its reading has been eliminated. That clock will simply remain at its position which was predicted from the previous time point. Since the most common type of error is a time step (a clock change which persists into the future), an administrative decision was made to correct the time state of a clock with a detected error so that the difference between the time of the reference clock and the clock with an error agreed with the actual observation. If the error was a read error which does not persist into the future, an error will probably be detected at the next time point which will have opposite sign and a reverse correction applied.

The remaining problem is the possibility of a frequency step. If a very small frequency step occurs, the Kalman algorithm has the flexibility to slowly adjust the estimated frequency of the clock to the new value. If the frequency

step causes a time change large enough to be detected, it is not known in one interval whether this is a time error or frequency step. In case a frequency step did occur, it is desirable to allow the frequency state to adjust quickly to the new value. In order to accomplish this, the diagonal element of the state covariance matrix, $P(t+1|t+1)$ corresponding to $y(t+1)$ of the clock in question is increased. The value that is added to this diagonal element is the minimum of $(2^*c/\delta(t))^2$ and $\delta(t)10^4$, where c is the clock time correction. This allows the recursion to adjust to a new frequency within a few time steps. Whether or not a frequency step occurred, the P matrix returns to its previous order of magnitude within a few time steps.

The estimate of "true" time is obtained from the state vector $X(t|t)$ which contains the estimates of $x(t)$ for each clock. The estimated values of x (t) for each clock are in relative agreement since their differences agree with the observed differences at time t. A relatively crude clock within the computer is used to determine the approximate time of the measurement.

Each of the clocks runs freely and independently of the others. Thus the ticks of the clocks each second are randomly distributed. If the tick on the reference clock is used as the indicator to read the other clocks, the readings are the time differences between the ticks in nanoseconds. At some time in the past the ensemble was initialized relative to some definition of "true" time (for example, international Atomic Time) and the uncertainty expressed in the initial covariance matrix $P(0|0)$. At that time each clock's tick was a certain number of nanoseconds off of "true" time. As data are collected and the algorithm progresses through its steps, all the clocks remain in relative agreement, but the ensemble drifts as a whole. The variance of this ensemble drift is given by the diagonal elements of $P(t|t)$ corresponding to the $x(t)$ for each clock. Since observational error is so small, these diagonal elements are essentially all the same. True time is estimated by correcting the tick of any one of the clocks by an amount equal to its estimated time state, $x(t)$, which is an estimate of the clock's deviation from true time. The variance of time estimate is the corresponding diagonal element of $P(t|t)$.

Sample output from the algorithm starting with the initialization, which was output from the previous run, is shown in figure 1. An example of a detected error with an adjustment is shown in figure 2. The algorithm also has the capability to adjust any clock administratively by a given amount or apply a frequency adjustment to all clocks in order to steer a time scale. These are software adjustments which do not affect the physical clocks but enable a time scale to trace another time scale such as International Atomic Time. The National Bureau of Standards has several of these paper scales.

CLOCK	INITIAL STATE TIME	FREQ	DRIFT	STANDARD DEVIATION TIME	FREQ	DRIFT	ROOT Q TIME	FREQ	DRIFT
1 1316	270534 6	9. 20	0. 000	1356. 2	5. 02	0. 000	4. 14	0. 80	0. 000
2 167	-44523. 9	349. 80	0. 000	1356. 2	6. 72	0. 000	13. 52	1. 11	0. 000
3 137	11731 0	52. 97	0. 000	1356. 2	7. 29	0. 000	11. 31	2. 49	0. 000
4 113	-207978285 4	939. 05	0. 000	1356. 2	7. 60	0. 000	9. 48	3. 18	0. 000
5 PHM	-254618838 5	660. 64	0. 000	1356. 2	4. 81	0. 000	0. 65	0. 77	0. 000
6 352	-78346 9	568. 72	0. 000	1356. 2	7. 37	0. 000	8. 85	3. 32	0. 000
11 1375	3038 1	-96. 37	0. 000	1356. 2	6. 20	0. 000	10. 71	1. 48	0. 000
14 601	12014 9	41. 48	0. 000	1356. 2	4. 61	0. 000	2. 13	0. 06	0. 000
15 0	-25018 5	-966. 53	0. 000	1356. 2	7. 05	0. 000	8. 65	2. 76	0. 000

MJD= 45060   TIME= 1200

DIFF	OBSERVATION	PREDICTED OBS	RESIDUAL	STANDARD DEVIATION	T TEST
1- 2	314518. 6	314517. 9	0. 7	13. 1	0. 0
1- 3	258754. 6	258759. 8	-5. 0	13. 4	-0. 4
1- 4	208247878 0	208247890. 2	-12. 2	12. 1	-1. 0
1- 5	254888722 0	254888721. 7	0. 3	4. 8	0. 1
1- 6	348332 6	348322. 0	10. 6	11. 6	0. 9
1-11	267581 0	267602 1	-21. 1	12. 4	-1. 7
1-14	258488 3	258487. 3	1. 0	5. 1	0. 2
1-15	296546 0	296528 8	17. 2	11. 1	1. 5

CLOCK	UPDATED STATE TIME	FREQ	DRIFT	STATE CHANGES TIME	FREQ	DRIFT	STANDARD DEVIATIONS TIME	FREQ	DRIFT
1 1316	270544 8	9. 31	0. 000	0. 9	0. 11	0. 000	1360. 7	5. 02	0. 000
2 167	-43973. 8	349. 80	0. 000	0. 2	0. 00	0. 000	1360. 7	6. 61	0. 000
3 137	11790 0	54. 08	0. 000	6. 0	1. 11	0. 000	1360. 7	7. 29	0. 000
4 113	-207977333 2	942. 67	0. 000	13. 1	3. 62	0. 000	1360. 7	7. 39	0. 000
5 PHM	-254618177 2	660. 74	0. 000	0. 6	0. 10	0. 000	1360. 7	4. 81	0. 000
6 352	-77787 8	565. 58	0. 000	-9. 7	-3. 14	0. 000	1360. 7	7. 37	0. 000
11 1375	2963 8	-93. 56	0. 000	22. 0	2. 80	0. 000	1360. 7	6. 20	0. 000
14 601	12056 3	41. 47	0. 000	-0. 1	0. 00	0. 000	1360. 7	4. 61	0. 000
15 0	-26001 2	-971. 05	0. 000	-16. 2	-4. 52	0. 000	1360. 7	7. 05	0. 000

FIGURE 1. Example of output of algorithm starting with the initialization from the previous run. Under INITIALIZATION and UPDATED STATE, the standard deviations are the square roots of the diagonal elements of the state covariance matrix, $P(t|t)$. Under $MJD = 45060$ the standard deviations are the square roots of the diagonal elements of the innovation covariance matrix, $C(t)$.

CLOCK	UPDATED STATE TIME	FREQ	DRIFT	STATE CHANGES TIME	FREQ	DRIFT	STANDARD DEVIATIONS TIME	FREQ	DRIFT
1 1316	270698 8	9. 80	0. 000	1. 3	0. 29	0. 000	1477. 2	5. 03	0. 000
2 167	-29561 0	355. 26	0. 000	-13. 8	-1. 07	0. 000	1477. 2	6. 09	0. 000
3 137	13186 1	58. 85	0. 000	14. 7	2. 97	0. 000	1477. 2	7. 29	0. 000
4 113	-207952798 7	939. 05	0. 000	0. 9	0. 38	0. 000	1477. 2	7. 39	0. 000
5 PHM	-254600975 3	662. 08	0. 000	-1. 8	-0. 68	0. 000	1477. 2	4. 82	0. 000
6 352	-63184. 8	560. 60	0. 000	9. 0	2. 94	0. 000	1477. 2	7. 37	0. 000
11 1375	417 2	-96. 96	0. 000	-9. 3	-1. 16	0. 000	1477. 2	6. 21	0. 000
14 601	13139 5	41. 47	0. 000	0. 0	0. 00	0. 000	1477. 2	4. 62	0. 000
15 8	-51091. 2	-965. 30	0. 000	6. 1	1. 79	0. 000	1477. 2	7. 06	0. 000

MJD= 45087   TIME= 1409

DIFF	OBSERVATION	PREDICTED OBS	RESIDUAL	STANDARD DEVIATION	T TEST
1- 2	299646 2	299642. 6	3. 6	14. 9	0. 2
1- 3	257496 6	257455. 0	41. 6	13. 7	3. 0
1- 4	208222466 2	208222483. 2	-17. 0	12. 4	-1. 4
1- 5	254870967. 7	254870961. 0	6. 7	4. 9	1. 4
1- 6	333293 3	333280. 6	-23. 1	11. 9	-2. 1
1-11	270444. 9	270393. 3	51. 6	12. 5	4. 1
1-14	257535 0	257524. 5	10. 5	5. 2	2. 0
1-15	322941. 2	322943. 4	-4. 2	11. 4	-0. 4

++++++QUAD=	40 6	CLOCK 11	-44. 7	11. 7	3. 8

DIFF	OBSERVATION	PREDICTED OBS	RESIDUAL	STANDARD DEVIATION	T TEST
1- 2	299646 2	299642. 6	3. 6	14. 8	0. 2
1- 3	257496. 6	257455. 0	41. 6	13. 7	3. 0
1- 4	208222466 2	208222483. 2	-17. 0	12. 4	-1. 4
1- 5	254870967. 7	254870961. 0	6. 7	4. 9	1. 4
1- 6	333293 3	333280 6	-23. 1	11. 9	-2. 1
1-14	257535 0	257524. 5	10. 5	5. 2	2. 0
1-15	322941. 2	322943. 4	-4. 2	11. 4	-0. 4

CLOCK 11 ADVANCED BY   -44. 7

CLOCK	UPDATED STATE TIME	FREQ	DRIFT	STATE CHANGES TIME	FREQ	DRIFT	STANDARD DEVIATIONS TIME	FREQ	DRIFT
1 1316	270715 7	7. 39	0. 000	10. 6	1. 58	0. 000	1482. 1	5. 02	0. 000
2 167	-28950. 9	355. 74	0. 000	7. 1	0. 48	0. 000	1482. 1	6. 07	0. 000
3 137	13219 1	51. 98	0. 000	-31. 0	-6. 87	0. 000	1482. 1	7. 23	0. 000
4 113	-207951750. 5	946. 44	0. 000	27. 7	7. 39	0. 000	1482. 1	7. 50	0. 000
5 PHM	-254600292 0	662. 99	0. 000	3. 9	0. 91	0. 000	1482. 1	4. 81	0. 000
6 352	-62309. 8	571. 19	0. 000	35. 6	10. 59	0. 000	1482. 1	7. 48	0. 000
11 1375	270. 8	-96. 70	0. 000	3. 7	0. 26	0. 000	1482. 1	99. 55	0. 000
14 601	13180 8	41. 45	0. 000	0. 1	-0. 02	0. 000	1482. 1	4. 62	0. 000
15 8	-52125 5	-961. 81	0. 000	14. 8	3. 49	0. 000	1482. 1	6. 99	0. 000

FIGURE 2. Example of computer output with an error detected in clock 11. The listing starts with the final results from the previous day. QUAD refers to the overall test which is not actually used for detecting errors but is printed out so that errors are easily visible. The optimal estimate of the clock error is -44.7 with standard deviation 11.7.

21

5. Estimation of Parameters

Assuming Gaussian errors, -2 ln likelihood is calculated from

$$L = \sum_t \{ln|C(t)| + I'(t)\, C^{-1}(t)\, I(t)\}. \qquad (5.1)$$

This gives a number which depends on the values of the parameters of the model. Nonlinear optimization routines can then be used to minimize this function giving maximum likelihood estimates of the parameters. Each function evaluation is a pass of the recursion through all the data.

When error detection is used, a new problem arises. The dimensions of $C(t)$ and $I(t)$ are reduced if a clock is eliminated at time t. When the nonlinear optimization routine varies the parameters, it is possible that at some point a clock that was previously eliminated is suddenly not eliminated, or that a clock that was not eliminated *is* eliminated. This would cause a discontinuity in the likelihood function. To avoid this problem, a file of indicator variables, indicating which clocks are deleted at which time points, is produced. The nonlinear optimization is then run conditional on fixed clocks being eliminated at certain time points. Upon convergence the parameters in the original model are replaced by their optimized value and the initial program rerun generating a new array of indicator variables. If the value of the likelihood function changes it indicates that the error detection procedure has detected a different set of errors so the nonlinear optimization is repeated. Experience has shown this procedure converges in two or three iterations.

The optimization was carried out on one year of data, April 1, 1981 to March 31, 1982. During this time 12 clocks were used, but since clocks were added and deleted administratively from the algorithm (removed because of problems with the clock, not error detection) a maximum of 10 clocks and a minimum of eight clocks were available at any one time. Figure 3 shows the times at which various clocks were present and the number of actual observations used for each clock. The number missing is the number of observations eliminated by error detection.

In the first nonlinear optimization run, σ_ξ and σ_η were varied for each clock and $w(t)$ and σ_a were set equal to zero. This is a drift-free model, and with 12 clocks gives 24 parameters to be estimated. Standard deviations rather than variances are used for optimization since if they go negative their squares are still positive. If variances were used directly, it would be possible for the nonlinear optimization routine to try negative values which would be meaningless. The results of this optimization procedure are shown in table 1.
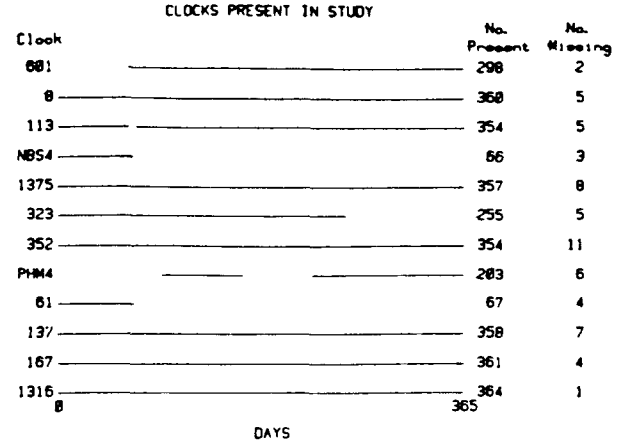

FIGURE 3. A schematic diagram showing the time when various clocks are available for use in the algorithm. The number present are the actual number used after error detection. The numbers missing are the number of errors detected. The sum of the number present and number missing is the number of observations available to the algorithm.

TABLE 1. *Estimated values of σ_ε and σ_η and 95% confidence intervals.*

Clock	n	$\sigma_\varepsilon \sim$ nanoseconds			$\sigma_\eta \sim$ nanoseconds/day		
		Lower Limit	Est	Upper Limit	Lower Limit	Est	Upper Limit
1316	364	3.81	4.14	4.53	0.53	0.80	1.23
167	361	12.58	13.52	14.67	0.57	1.11	2.07
137	358	10.41	11.31	12.27	1.76	2.49	3.56
61	67	5.48	6.77	8.43	1.53	2.80	4.83
352	354	8.12	8.85	9.74	2.42	3.32	4.41
323	255	2.06	2.37	2.74	0.63	0.94	1.34
1375	357	9.93	10.71	11.64	0.96	1.48	2.25
NBS4	66	0	0.88	1.86	0.72	1.34	2.16
113	354	8.73	9.48	10.38	2.49	3.18	4.11
8	360	7.98	8.65	9.49	2.11	2.76	3.66
601	298	1.89	2.13	2.41	0	0.06	0.52
PHM4	203	0	0.65	1.19	0.55	0.77	1.09

When m clocks are in the model, it is only possible to estimate m-1 drift terms when the observations are time differences. Tryon and Jones [1] handled this problem by estimating m drifts with the constraint that the sum of the drifts was zero. Here, a different approach is used. A clock in the model is assumed to have zero drift and the drifts of the m-1 remaining drifts are estimated. Clock 601 was chosen to be the clock with no drift since the standard deviation of its frequency random walk was not significantly different from zero. Since clock 601 was not available for the entire year, the drift for clock 323 was also set equal to zero. These two clocks had overlapping data intervals which included the entire year. Clock 323 was also chosen because it is known to be a particularly

22

stable clock. The drifts for two other clocks, 61 and NBS4, were also set equal to zero because of their short lengths of data. The remaining eight clocks were tested for drift by a 24-parameter nonlinear optimization run. For each of the eight clocks, the frequency random walk standard deviation, σ_η, the initial value of the drift, $w(0)$, and the drift random walk standard deviation, σ_α, were allowed to vary. All of the time random walk standard deviations, σ_ξ, were fixed at their optimized values from the previous run, as were the parameters for the remaining four clocks in the algorithm. This approach of fixing some of the parameters was necessitated by the storage available on the PDP 11/70 which allowed a maximum of 25 parameters to be optimized simultaneously. Changes in σ_ξ that occur when drift is included in the model, while small, do in fact exist [1].

The estimated values of all eight σ_α were very small, so their significance was tested by fixing their values at zero and rerunning the previous optimization with 16 parameters. Statistical tests based on changes in $-2 \ln$ likelihood have asymptotic chi-square distributions with degrees of freedom equal to the number of parameters added or deleted from the model or constraints introduced [2]. In this case, eight parameters were fixed at zero but $-2 \ln$ likelihood only increased by 0.2 indicating that the σ_α are not significantly different from zero. This agrees with the results obtained on the earlier data set [1]. These results, which were obtained on a CDC Cyber 170/750 computer, were checked using the new algorithm on the PDP 11/70 and duplicated exactly. At this point it is assumed that any frequency drifts which may exist can be considered to be constant over the length of the data span.

When the eight drift terms were eliminated from the model, the change in $-2 \ln$ likelihood was not significant, indicating that the inclusion of drift terms did not improve the model significantly. This did not agree with the previous study where there were significant trends for clocks 601, 137, and 61, but there are explanations for this. In the first study clock 601 was nearing the end of its life, and the cesium beam tube was replaced between the two studies making it essentially a different clock. Clock 137 was introduced as a new clock on the first day of the first study. Clocks are known to drift near the beginning and end of their lives and be more stable during midlife. Clock 61 was only available for 69 days in the second study so any drift that may have existed could not be detected. If a common drift exists in all clocks, this could not be estimated, and the uncertainty introduced into the time scale would not appear in the $P(t|t)$ matrix.

It is of interest to compare the confidence intervals obtained on two different sets of data. Table 2 shows the results for the clocks that were common to the two studies. The estimates and confidence intervals are in agreement except for clock 601 in which the cesium beam tube was replaced.

TABLE 2. *1979 and 1981 clock characterization 95% confidence intervals.*

Clock	Study	$\sigma_\eta \sim$ nanoseconds			$\sigma_\alpha \sim$ nanoseconds/day		
		Lower Limit	Est	Upper Limit	Lower Limit	Est	Upper Limit
601	1	6.88	7.46	8.14	0	0.44	0.96
	2	1.89	2.13	2.41	0	0.06	0.52
167	1	12.44	13.45	14.61	0.56	1.11	1.96
	2	12.58	13.52	14.67	0.57	1.11	2.07
137	1	9.26	10.04	10.96	1.03	1.60	2.26
	2	10.41	11.31	12.27	1.76	2.49	3.56
1316	1	3.17	3.62	4.05	0.97	1.36	1.82
	2	3.81	4.14	4.53	0.53	0.80	1.23
323	1	3.11	3.53	3.94	0.41	0.73	1.10
	2	2.06	2.37	2.74	0.63	0.94	1.34
8	1	8.32	9.09	9.95	2.01	2.65	3.48
	2	7.98	8.65	9.49	2.11	2.76	3.66

6. Conclusion

A major advantage of a Kalman algorithm for estimating time is its flexibility when changes are needed. An example is the introduction of frequency calibrations from a frequency standard. A frequency calibration, together with its precision, can easily be introduced into the state vector and state covariance matrix. Experimentation with different Kalman models are possible since changing the state transition matrix is not difficult. However, the major advantage is that the Kalman algorithm is based on a mathematical model which fits the data. The innovations can be tested for whiteness, and standard statistical methods can be used for obtaining confidence intervals and hypothesis tests.

Acknowledgment

References

[1] Tryon, P. V. and Jones, R. H. (1982). Estimation of parameters in models for cesium beam atomic clocks. Second Symposium on Atomic Time Scale Algorithms, June 23-25, 1982, Boulder, CO.

[2] Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the theory of statistics*. Third edition. New York: McGraw-Hill.

[3] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME (Journal of Basic Engineering), 82D, 35-45.

[4] Graybill, F. A. (1976). *Theory and application of the linear model*. North Scituate: Duxbury Press.

Appendix

For computational efficiency and numerical stability, a procedure involving a Cholesky decomposition is used in the Kalman recursion [4]. The tests for errors are also simplified. The matrix product $H(t+1)\ P(t+1|t)$ in eq. (4.5) is calculated without actually performing a matrix multiplication. Since each row of $H(t+1)$ contains a one and a minus one with the rest of the elements zero, each element of the matrix product is simply the difference between two elements of $P(t+1|t)$. The matrix $C(t+1)$ is calculated and augmented as follows

$$[C(t+1) \vdots H(t+1)\ P(t+1|t) \vdots I(t+1) \vdots A_1 \vdots A_2 \vdots \ldots \vdots A_m]. \quad (A.1)$$
$$(m-1) \qquad 3m \qquad\qquad 1 \quad 1 \quad 1 \quad \cdots \quad 1$$

This matrix has $m-1$ rows with the number of columns shown under each partition. Only the upper triangular position of $C(t+1)$ need be stored. The third partition is the vector of innovations from eq (4.4) followed by the vectors to be used for testing for clock errors. One of these vectors will correspond to the reference clock eq (4.7) and the others to the remaining clocks eq (4.9).

The Cholesky decomposition factors $C(t+1)$ into

$$C(t+1) = T'(t+1)\ T(t+1), \quad (A.2)$$

where $T(t+1)$ is upper triangular. The algorithm as given in [4] works on the entire augmented matrix and is equivalent to premultiplying the matrix by $[T'(t+1)]^{-1}$. Let

$$B(t+1) = [T'(t+1)]^{-1}\ H(t+1)\ P(t+1|t)$$

$$D(T+1) = [T'(t+1)]^{-1}\ I(t+1) \quad (A.3)$$

$$G_i = [T'(t+1)]^{-1}\ A_i.$$

The test for each clock, eq (4.10) becomes

$$b = G'D(t+1)/G'G, \quad (A.4)$$

and eq (4.11) becomes

$$s.e.(b) = (1/G'G)^{\frac{1}{2}} \quad (A.5)$$

The quadratic form in eq (4.13) reduces to

$$Q = D'(t+1)\ D(t+1), \quad (A.6)$$

and eq (4.15) to

$$\Delta(t+1) = B'(t+1)\ [T'(t+1)]^{-1}. \quad (A.7)$$

The final updating equations are (4.16)

$$X(t+1|t+1) = X(t+1|t) + B'(t+1)\ D(t+1), \quad (A.8)$$

and (4.17)

$$P(t+1|t+1) = P(t+1|t) - B'(t+1)\ B(t+1). \quad (A.9)$$